Designing a Predictive Coding System for Electronic Discovery

Dhivya Soundarajan (M.S. student, Human-Computer Interaction) and Sara Anne Hook, M.B.A., J.D.
Department of Human-Centered Computing
School of Informatics and Computing
Indiana University-Purdue University Indianapolis

Not long ago, the concept of using predictive coding and other technologies to assist with the electronic discovery process seemed revolutionary. *Da Silva Moore* and *Global Aerospace* stand as the first major cases where judges strongly supported predictive coding.[1-2] A recent Indiana case recognized it as a useful method for reducing the amount of potentially relevant evidence that has to be searched and culled.[3] Within just a few short years, using predictive coding as part of an electronic discovery process is now considered acceptable and perhaps even expected. It is not difficult to appreciate the advantages of predictive coding and its superiority over a manual process at various steps of electronic discovery, particularly during the review step.[4-11] However, questions still remain about the efficacy of the predictive coding process and the tools that are available.[12-13] Because the use of predictive coding systems in law is still in its infancy, it presents us with an opportunity to design something that will not only take advantage of the power of big data and computational algorithms, but that will also incorporate design and usability principles to provide an attractive and easy-to-use interface for lawyers to interact with. Predictive coding uses natural language processing and other mathematical models to enhance search results, but the essence of these systems is that they actually learn and the precision of the retrieval improves as additional collections of evidence are entered. Behind-the-scenes will be a repository where all of the evidence for a case resides. Our system will assist the lawyers in reducing the time and cost of an electronic discovery process as well as minimize the chances for mistakes in determining which evidence is relevant to a case and which evidence can be withheld under attorney-client privilege, as attorney work-product or another confidentiality doctrine.

1. *Da Silva Moore v. Publicis Groupe & MSL Group*, No. 11 Civ. 1279, 2012 WL 607412 (ALC) (AJP) (S.D.N.Y. Feb. 24, 2012).
2. *Global Aerospace, Inc. v. Landow Aviation, L.P.*, No. CL 61040 (Vir. Cir. Ct. Apr. 23, 2012).
3. *In re Biomet*, 2013 WL 1729682 (N.D. Ind. Apr. 18, 2013).
4. Alison Silverstein and Geoffrey Vance. E-Discovery Myth Busters: Why Predictive Coding is Safe, Successful and Smart. *Peer to Peer*, Vol. 29, No. 4, December 2013, pp. 66-69.
5. John Papageorge. Predictive Coding Gaining Support in Courts. *Indiana Lawyer*, January 29-February 11, 2014, p. 8.
6. Adam M. Acosta. Predictive Coding: The Beginning of a New E-Discovery Era. *Res Gestae*, October 2012, pp. 8-14.
7. Ajith (AJ) Samuel. Analytics Driving the E-Discovery Process. *Peer to Peer*, Vol. 28, No. 2, June 2012.
8. Richard Acello. Beyond Prediction: Technology-Assisted Review Enters the Lexicon. *ABA Journal*, August 2012, pp. 37, 70.

9. Barry Murphy. The Rise of Technology-Assisted Review (TAR). *Peer to Peer*, Vol. 28, No. 2, June 2012, pp.
10. Brian Ingram. Controlling E-Discovery Costs in a Big Data World. *Peer to Peer*, Vol. 29, No. 1, March 2013.
11. Hal Marcus and Susan Stone. *Beyond Predictive Coding - The True Power of Data Analytics* [webinar]. International Legal Technology Association, May 19, 2015.
12. Jessica Watts and Gareth Evans. *Predictive Coding in the Real World* [webinar]. International Legal Technology Association, August 5, 2015.
13. Danielle Bethea. Predictive Coding: Revolutionizing Review or Still Gaining Momentum? *Litigation and Practice Support: ITLA White Paper*, International Legal Technology Association, June 2014.